

REMARKS

Reconsideration and withdrawal of all grounds of rejection are respectfully requested in view of the above amendments and the following remarks. Claims 1-15 were rejected. By this Amendment, claim 7 has been amended. No claims have been cancelled. Consequently, claims 1-15 are now pending.

The Examiner has objected to the title of the application because it is not descriptive. Specifically, the title is not clearly indicative of the invention to which the claims are directed. Appropriate correction to the title is included in this Amendment. Accordingly, it is requested that this objection be withdrawn.

A minor grammatical error in a paragraph beginning on page 2, line 10 has been corrected. Entry of this change is respectfully requested.

The Examiner has rejected claims 1-6 under 35 U.S.C. § 102(b) as being anticipated by U.S. Patent No. 6,012,058 to Fayyad et al. The '058 patent to Fayyad et al. is generally directed to an apparatus and method for clustering data into data sets that characterize the data. More specifically, Fayyad et al. teaches the use of a data mining system which utilizes scalable clustering techniques to evaluate data within large databases. As the Examiner will note below, the undersigned was well-aware of the '058 patent to Fayyad et al. when preparing the present application. (see page 2, lines 2-7 of the application as filed).

United States Patent No. 6,012,058 to Fayyad et al. which issued January 4, 2000 discloses one data mining process for clustering data. The disclosure of this patent is incorporated herein by reference. This patent discloses a clustering process that extracts sufficient statistics concerning a large database to produce a data clustering model that takes up far less memory than the entire database.

Many sections of the '058 patent cited by the Examiner refer directly to Figure 4 of the patent. As disclosed, Figure 4 is a flow diagram of an exemplary embodiment of the invention. The process teaches generally involves the following sequential steps:

Step 1:	Initialize Data	Ref. Char. 100
Step 2:	Get Data	Ref. Char. 110
Step 3:	Do Extended K-Means Clustering	Ref. Char. 120
Step 4:	Calculate RS, DS, CS	Ref. Char. 130
Step 5:	Terminate/Suspend Decision	Ref. Char. 140

Step 1 is an initialization step 100 and includes initializing a number of data structures and choosing a cluster number K, i.e., the total number of clusters within the data, for characterizing the data. In a second step 110, a portion of the data in the database 10 is sampled and brought within the random access memory of a computer. A processor unit or similar device next executes step 3, a clustering procedure 120 using the data brought into memory in the Step 2. The processor 21 assigns data contained within the portion of data brought into memory to a cluster and determines a mean for each attribute of the data assigned to a given cluster. A data structure for the results or output model of the analysis is a model of the clustering of data. As taught by the '058 patent, each record has three required components sufficient to compute the mean and covariance of the data in a cluster. Step 4 is a summarization step 130 of at least some of the data used in the present iteration to characterize the K clusters. Before potentially looping back to get more data, Step 5 is performed. The processor 21 determines 140 whether a stopping criteria has been reached. If not, Steps 2-4 are repeated.

The present invention provides a new and improved method to quickly identify a random subset of the data within a large database and run a data mining algorithm on the subset instead of the whole data set. The total time needed to model the data is much smaller than the time it would take to run the analysis on the original (large) data set. The invention seeks to solve an identified need in the art as discussed on page 1, line 26 to page 2, line 2.

Although many algorithms for such problems are known and widely used (for example, Decision Trees and K-Means Clustering), they take too much time if trained on too much data. It has been observed that under certain circumstances, however, it may not be necessary to use an entire database (which can have many millions of records) to create a useful model or predictor. Instead a sample of a few tens of thousand records might accurately represent the much larger data set of the entire database.

Independent claim 1 of the present invention is directed to a computer-implemented process. As pending, claim 1 reads as follows:

1. A method of identifying a subset of records within a database for purposes of representing said database comprising:
 - a) choosing a selection attribute from one of a plurality of attributes contained in records within the database;
 - b) scanning records in the database and applying a randomizing function to the selection attribute of each record to create a randomized record value; and
 - c) applying a selection criteria to identify records for inclusion within a subset of records by comparing the randomized record value of each record with the selection criteria.

In a rejection of the first above recited method step, the Examiner has cited col. 4, lines 30-34 and 43-60. The cited passages teach two methods steps 100, 110 previously discussed above. As cited, Fayyad et al. does not teach or suggest choosing a selection attribute for a plurality of attributes contained in records within the database. The disclosure cited simply discusses pre-clustering and clustering techniques that involve testing *every* attribute of *every* data record. Any material mention of attributes in the cited section is in regard to clustering, i.e., “the dimension of the vectors is the number of attributes of the data records in the data base” (‘058, col. 4, lines 56-58) and not in regard to selecting an attribute for data mining. Therefore, this limitation is clearly not anticipated by Fayyad et al. ①

The Examiner has cited two sections, col. 4, line 61 to col. 5, line 6 and col. 13, lines 5-24, in rejecting the second recited method step of claim 1. The first cited passage contains references to “random sampling of data” performed in Step 2 of the clustering process. This and other types of random data sampling is unrelated to the randomizing function recited in claim 1. ② The second cited passage provides additional detail about specific clustering techniques. The second element of claim 1 is directed to applying a randomizing function to the specific attribute that was selected in the first method step of claim 1. The process results in a randomized record for each set of records scanned. As cited, Fayyad et al. does not teach or suggest the use of a randomizing function applied to a section attribute to produce a randomized record for a dataset. It is respectfully submitted that the two cited sections of Fayyad et al. are related art at best. In view of the above discussion, the second element of claim 2 limitation is clearly not anticipated by Fayyad et al.

In a rejection of the third recited method step above, the Examiner has cited col. 9, lines

20-45 of the '058 patent. Admittedly, the cited passage does teach a clustering technique that involves a selection process. However, the process is in regard to cluster selection and not in regard to individual records selection. Specifically, the cited passage teaches a method to select "dense clusters" by eliminating clusters that are "either too small (in terms of number of points) or too 'spread out'." Fayyad et al. clearly does not teach, suggest or disclose applying a selection criteria to identify records for inclusion within a subset of records by comparing the randomized record value of each record with the selection criteria. Therefore, the third element of claim 1 is clearly not anticipated by Fayyad et al.

For at least the reasons set forth above, it is respectfully submitted that the '058 patent does not anticipate claim 1 of the present application. Furthermore, there is no suggestion of how the Fayyad et al process would be modified to meet the recitations of claim 1. As a result, claim 1 is allowable. Further, it is submitted that claims 2-6 are allowable at least by virtue of direct or indirect dependence on allowable claim 1. Consequently, withdrawal of this rejection is respectfully requested.

The Examiner has rejected claims 7-9 under 35 U.S.C. § 102(b) as being anticipated by U.S. Patent No. 6,012,058 to Fayyad et al.

Independent claim 7 of the present invention is directed to a client/server computer apparatus. As amended, claim 7 reads as follows:

7. A Client/Server computer apparatus comprising:
one or more client computers coupled to a network and including communications instructions for requesting a data set by means of the network; and
a server computer coupled to the network and having access to a database having a number of records, said server computer including instructions for sending a dataset made up of a subset of the records in the database to a client computer via the network;
said server computer including instructions for scanning records in the database, applying a randomizing function to a specified record attribute of each record in the database to produce a randomized record value, and comparing the randomized record value with a selection criteria to determine whether to include a record in the subset of records from the database for transmission via the network to the client.

In rejecting the third element of claim 7, the Examiner has repeated the citation of the passage from col. 4, line 61 to col. 5, line 6 of Fayyad et al. For at least the reasons provided above in regard to the patentability of claim 1, Fayyad et al. does not anticipate this element of

claim 7. As a result, claim 7 is allowable. Further, it is submitted that claims 8-9 are allowable at least by virtue of dependence on allowable claim 7. Consequently, withdrawal of this rejection is respectfully requested.

The Examiner has rejected claims 10-15 under 35 U.S.C. § 102(b) as being anticipated by U.S. Patent No. 6,012,058 to Fayyad et al.

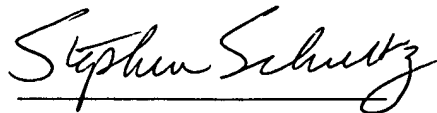
Independent claim 10 is directed to a computer readable medium including instructions for identifying a subset of records within a database for purposes of representing the database. In short, the claim recites instructions for performing the steps recited in claim 1. Therefore, for at least the same reasons as provided above in regard to the patentability of claim 1, Fayyad et al. does not anticipate claim 10 and consequently that claim is allowable. Further, it is submitted that claims 11-15 are allowable at least by virtue of dependence on allowable claim 10. Consequently, withdrawal of this rejection is respectfully requested.

In view of the above, it is respectfully submitted that the invention of independent claims 1, 7 and 10 is patentable. Further, the subject matter of the remaining dependent claims is patentable at least by virtue of direct or indirect dependence on claims 1, 7 and 10. Therefore, it is believed that all pending claims of this application are in condition for allowance. Accordingly, entry of the Amendment and a subsequent early Notice of Allowance for all pending claims of this application is respectfully solicited.

Respectfully submitted,

Date: Oct 2, 2003

Telephone: (216) 241-6700
Facsimile: (216) 241-8151



Stephen J. Schultz
Reg. No. 29,108